Testimony on "Following Toxic Clouds: Science and Assumptions in Plume Modeling"

Steven R. Hanna

2 June 2003

My name is Steven Hanna. I am an Adjunct Associate Professor at the Harvard School of Public Health, a Research Professor at George Mason University, and president of Hanna Consultants. For the past 38 years, my career has emphasized plume modeling, beginning with my graduate research in Meteorology at Penn State, continuing with 14 years' experience at NOAA's Atmospheric Turbulence and Dispersion Laboratory in Oak Ridge, TN, and followed by 18 years' experience with environmental consulting firms in Massachusetts. For the past six years, I have held concurrent appointments at Harvard and at George Mason University, and have carried out related research studies through my consulting company. Throughout this period I have developed and evaluated plume models and studied air pollution meteorology for a wide range of applications and for a mix of government and industrial sponsors. Because of this broad experience, I am often called upon to carry out independent assessments of plume models and have chaired several peer review committees for plume models such as the EPA's new AERMOD model. From 1997 through 2001, I chaired the peer review committee for the plume modeling done for the Gulf War. For nine years (1989-1998), I was the Chief Editor of the *Journal of Applied Meteorology*, where I made the final publication decision for over 1000 manuscripts.

Because of my broad experience in developing and evaluating emergency response plume models for industry and government over the past 20 years, I am familiar with the models and scenarios used by a wide array of groups. I am not linked with a particular model or group and can offer unbiased opinions on the models' capabilities and on their strengths and weaknesses.

Before answering the questions listed in my invitation letter, I wish to point out that plume models are sets of mathematical and/or computer equations that are used to estimate the location and magnitude of concentrations or dosages (concentrations summed over time) due to releases of contaminants to the atmosphere. Combined with information on health effects, the results of the plume model can be used to make emergency response decisions in real time. Other uses of plume models are to carry out planning exercises or determine the effects of a past incident. The type of contaminant (e.g., gas, particles, aerosol) does not usually matter to the calculations, since all dilute gases and small particles are transported and dispersed alike. In the case of large particles (i.e., diameters larger than about 100 micrometers) or large releases of dense gases (e.g., a rupture of a ten ton chlorine tank), there have been special plume formulas developed that are used as options in most hazardous gas models.

My answers to the 14 questions asked in my invitation letter are given below:

Question 1 – What are the types of dispersion models?

Dispersion models are applied when hazardous materials are emitted to the atmosphere at an assumed mass per unit time over an assumed period of time. All dispersion models are similar in that they calculate two basic characteristics of the emitted material – 1) the speed and direction that the plume moves with the wind, and 2) the dispersion or lateral and vertical spread due to turbulence in the atmosphere. Simple fast-running models called Gaussian plume models have been successfully applied for many decades. Examples are the EPA's ISC model and NOAA's CAMEO/ALOHA. Slight modifications to these simple models have been made to account for changes in wind speed and direction with time, and the resulting models are called Lagrangian puff models, which form the basis for applied modeling systems such as the EPA's CALPUFF, NOAA's HYSPLIT, and DTRA's HPAC/SCIPUFF models. A related model based on Lagrangian particles is DOE's NARAC system. In addition to the many government-sponsored models, there is a class of high-quality models developed by the chemical and oil industries for application to accidental releases of hazardous chemicals (e.g., hydrogen fluoride, chlorine, or propane) to the atmosphere, where algorithms are needed to handle high-velocity aerosol jets and dense gases. The above models are most commonly used for emergency response since they can be run relatively quickly. Another widely-used modeling system that requires much more time is a three-dimensional grid model such as the EPA's Models3/CMAQ, which is applied to urban and regional ozone, particle, and toxics problems. At the far end of the spectrum of complexity are the Computational Fluid Dynamics (CFD) models, which are often applied in research mode to groups of urban buildings and are based on small three dimensional grids and which can take hours or days to run on a large computer.

Question 2 – What are the types of emergency response models?

Several types of emergency response models are in use by different groups, but the fundamental requirements are that the model must be run quickly and easily. The three-dimensional grid models discussed under Question 1 do not satisfy these requirements. Examples of the more commonly-used emergency response models that do satisfy these requirements are the Lagrangian Gaussian puff models HPAC/SCIPUFF (from DTRA), HYSPLIT (from NOAA) and VLSTRACK (from the Navy), the Lagrangian particle model NARAC (from the DOE), and several Gaussian plume/puff models with dense gas capabilities (CAMEO/ALOHA from NOAA, HGSYSTEM from the chemical and oil companies, and SAFER/TRACE and PHAST from consulting companies servicing primarily the nuclear and chemical industries).

Note that, in addition to the standard "source modeling" mode of predicting concentrations or dosages from a given source, the plume models can also be operated in "reverse or receptor-modeling" mode if the source location and magnitude are not known. In the latter situation, observations of concentration or dosage at two or more positions can be combined with the plume model to triangulate a best guess of the position and magnitude of the source.

Question 3 – What are the strengths and weaknesses of these models?

The major strength of the models mentioned above is that they have all been evaluated and "calibrated" with the eight or ten major sets of data from field experiments so that we can be confident that they produce results that agree with observations within a factor of about two for some simple release scenarios. Models such as HPAC/SCIPUFF, HYSPLIT, and NARAC are based on up-to-date science. The models for the chemical industries are applicable to a broader range of release types, including dense gases and aerosol jets.

The major weakness of these models is that any real source release is nearly always more complicated than the simple scenarios studied in the field and wind tunnel experiments on which the models are based. Real sources tend to be variable in time and space, to occur in non-ideal locations such as next to a building near a river, and to occur at times when the atmosphere is variable or rapidly changing or it is raining. Consider Bhopal, Three Mile Island, Chernobyl, and Khamisiyah, which all took place during non-ideal meteorological conditions with poorly-known sources.

Question 4 – What are the deficiencies in emergency-response or real-time models?

Several deficiencies were mentioned in my answer to Question 3. In addition, there are problems due to a lack of nearby meteorological data such as wind speed and direction and stability. Usually the models are developed and tested using highly-instrumented field experiments. However, in a real emergency application, the only available wind speed may be from an airport 20 km away.

Another deficiency is related to the need to clearly communicate the uncertainty in the model predictions. HPAC/SCIPUFF is the only model that includes an estimate of the uncertainty along with its forecasts. However, it is important for decision-makers to know that, even in the best of conditions, the model predictions can be expected to be accurate only within a factor of about two. Also, any uncertainties in the source term are directly translated into uncertainty in the modeled concentrations or dosages.

A deficiency in the OFCM review of emergency response models is that the useful models developed by the oil and chemical industries were generally ignored just because they were not developed by government agencies. Models such as HGSYSTEM, developed by a consortium of industries such as Shell and ExxonMobil, are in the public domain and account for a much wider range of chemical plant accident scenarios and source types than the government models. These industry models generally also include models for source emissions, such as flashing jets from pressurized HF tanks, or evaporation from LNG spills.

There are problems in knowing which agency is responsible for applying plume models to be used for emergency response decisions in some scenarios. The lines are fuzzy and responsibilities not entirely clear.

Question 5 – How can these deficiencies be improved?

The models should be improved to cover a wider range of meteorological scenarios and source scenarios, including difficult topics such as a chemical agent release that varies in time over 30 minutes and position over 5 km (e.g., a release from a moving truck) during a morning rush-hour with the sun rising.

The expected uncertainties of models should be explained as part of the training, and the uncertainty of a specific emergency response prediction should be included in the prediction. For example, if the model predicts that the plume will move towards the east, the decision makers should know that there is a chance that the plume will move towards the west.

The government agencies and the oil and chemical industries should work together so that the government models make optimum use of the scientific developments by the industries. Furthermore, the various agencies with plume models should decide the mechanism whereby different agencies will take the lead for running their models for certain scenarios so that it is clear which agency is responsible. It then follows that the various models should give fairly consistent results so that there are not large differences in emergency response actions taken by different agencies to the same release scenario.

Question 6 – What sources of data are needed for effective plume modeling?

The details of the source (location, mass release rate, duration, temperature and composition) are needed, since the accuracy of the plume model is no better than the accuracy of the source inputs. Information on land-use and nearby complex terrain or buildings is useful. Wind and stability inputs are necessary and should be from an unobstructed location as close to the source as possible. In combat zones and in remote sites where observed winds are unlikely to be available, meteorological forecast models can be run to provide wind inputs, although actual observations are preferred if available. Locations of critical populations are needed so as the focus the predictions on specific areas. Sometimes remote sounders are available that can provide real-time observations of winds for input to the plume models or of concentrations for use in refining (calibrating) the predictions.

Question 7 – How are plume models tested and validated?

Over the past few decades, there have been several field and laboratory (e.g., wind tunnel) experiments where tracer gas or small particles are released at a known rate, winds and stability are measured on-site, and concentrations or dosage are measured by numerous sampling instruments. Some of the field experiments carried out by DOE and by industry involved releases of hazardous gases and aerosols such as HF and ammonia. Some of the DOD experiments involved "real" scenarios such as exploding bunkers but used tracer gases. The field experiments are preferred over the laboratory experiments because they more closely represent actual release scenarios. However, the field experiments are relatively expensive (several million dollars for a several day study with 10 to 20 release trials).

In most cases, it does not matter what is used as a tracer, since most contaminants in the atmosphere are transported and diffused in the same manner. Most of the recent experiments used SF_6 as a tracer. Because of the costs and the difficulties, these experiments usually produce limited amounts of data and are carried out under relatively ideal conditions. If it rains or if a front goes through, the experimentalists pack up and go home. The total data archive of useful field experiments consists of about 20 or 30 experimental campaigns with about 10 to 50 individual trials per campaign. I have evaluated many models with most of these available data sets and am one of the few persons who have compared models from a variety of agencies and industries with the same data sets.

The model outputs that are evaluated depend on the needs of the decision makers. For example, sometimes all that is needed is the maximum distance to which a toxic concentration extends, while other times the need is for the precise spatial coverage of toxic concentrations. The models are generally evaluated by looking at the relative mean bias (e.g., the model overpredicts by 70 % on average) and the relative scatter (e.g., an individual prediction could be up to a factor of three high or low). We often determine whether differences between models are significant. For example, when we evaluated the EPA's CALPUFF model versus DTRA's HPAC model versus the Navy's VLSTRACK model with the Dipole Pride 26 field data from the Nevada Test Site, we found that all three models performed fairly well (i.e., relative mean biases less than a factor of two and relative scatters less than about a factor of 3 or 4). In fact because of natural variability in the atmosphere, it is unlikely to have a model that consistently performs better than this.

Question 8 – What are the challenges for developing sound models for cities?

Cities obviously are unique and every building is different and has its own special shape, roof structures, porches, nearby landscaping, etc. However, it is impractical in emergency response models to try to model the effects of each windowsill and HVAC system. The current approach by many agencies is to develop simplified three-dimensional models for the areas right around individual buildings or in specific street canyons, and then use simplified "urban canopy" models for the plume after it is transported past the first two or three buildings. We find that the urban boundary layer is well-mixed due to the effects of the buildings, making it possible to derive straightforward general relations. These relations are now being developed and tested using recent field experiments in Salt Lake City, Los Angeles, and San Diego. An extensive tracer experiment is planned in Oklahoma City for this July and should help refine and test these models.

Our recent evaluations of HPAC with the Salt Lake City data used several optional wind inputs, covering the range from only the airport winds to the complete set of special wind observations in the city, with the surprising result that the model performance was not best for optimum wind inputs. Further study of this issue is needed since there may be compensating errors.

Question 9 – What models should be used in the event of a chemical, biological, or radioactive release into the air?

As stated in paragraph three of my introduction, the same models can be used for chemical, biological, and radioactive releases, since all of these substances disperse in the atmosphere in the same manner. The most widely-used real-time models used for emergency response are the DOE NARAC model, the DTRA HPAC/SCIPUFF model, the NOAA HYSPLIT and CAMEO/ALOHA models, and several models used in the chemical industries (e.g., SAFER, PHAST, CHARM, HGSYSTEM). The EPA's AERMOD and CALPUFF models are excellent state-of-the-art models that could be used for emergency response, although they are currently used primarily for regulatory applications.

Question 10 – What is the margin of error for these models?

The plume models listed under Question 9 have similar performance measures (i.e., margins of error), and further improvement is probably limited by natural uncertainties in the atmosphere – the same uncertainties that cause weather forecasts to never be exactly correct.

The margin of error is less for well-defined scenarios and is more for poorly-defined scenarios. For field experiments where the source is well known and there is extensive on-site meteorology, the relative mean bias for a good model is less than about 30 or 40 % and the relative scatter is about 100 %. For real-world emergency response scenarios, the relative mean bias would be about a factor of two or three and the relative scatter would be about a factor of five to ten. However, this uncertainty of the plume model is still less than the uncertainty for the emissions rate or for the exposure and risk components of the model.

Question 11 – How were possible chemical warfare agent releases modeled in determining potential exposures in the 1991 Persian Gulf War?

From 1997 through 2001, I was the chair of the Peer Review Committee for the plume modeling of Khamisiyah. The plume modeling procedure was hampered from the start due to a deficiency in information on the source emissions rate and an almost complete lack of meteorological data in Iraq. The modeling methodology changed several times during this period as models were upgraded. For example, the OMEGA (DTRA) and COAMPS (Navy) mesoscale meteorological model went through several modified versions over the course of the study. The MM5 meteorological model was also applied in order to generate wind fields over the modeling domain. These meteorological models had to be used because of the lack of wind observations. Two plume models (DTRA's HPAC/SCIPUFF and the Navy's VLSTRACK) were applied, using best estimates of the source emissions and using the outputs of the mesoscale meteorological models. In order to account for the uncertainty, the researchers ran the two plume models separately using the wind inputs from the three meteorological models. The end product of each meteorological model and plume model combination was a map containing a set of dosage contours for three or four dosage limits based on health effects. An advisory committee decided to use the total area coverage of the contours predicted by the six model combinations (three meteorological models and two plume models) in order to define the area of possible health effects for U.S. troops.

Question 12 – What were the strengths/weaknesses of these models?

It is hard to tell what the strengths and weaknesses of the models were because there were no data whatsoever to use for comparisons. The Kuwait oil fire observations were of little relevance because those plumes rose to a much higher elevation. However, it was obvious that the three meteorological models often produced differences in wind direction of 30 or more degrees and differences in wind speeds of a factor of two, causing the plume paths to often diverge by 30 or more degrees and the cloud speeds to diverge significantly. The two plume models produced further differences. Our opinion is that the HPAC/SCIPUFF plume algorithms are slightly more state-of-the-art than VLSTRACK, although VLSTRACK has an excellent data base for CBN source emissions scenarios

Question 13 – What models should DOD use today should an event occur in a combat theater?

My answer has two parts. First, of the available DOD models, HPAC/SCIPUFF is probably the recommended model because of its state-of-the-science algorithms, its good performance against available field data, its useful graphical outputs, and its implementation within the context of DOD modeling centers and procedures. Second, there are other models available from other agencies and industries that are just as good technically and perhaps better in some areas, but which are not specifically formatted for DOD use. These models include the EPA's AERMOD and CALPUFF systems, the DOE's NARAC system, NOAA's HYSPLIT system, and industrial models such as HGSYSTEM. Even the simple NOAA CAMEO/ALOHA modeling system has been shown to agree well with observations, and includes technical details such as dense gases. In addition, the European Union, Australia, and other non-US countries have excellent models available such as the UK ADMS model. Thus I recommend that DOD review and consider making use of the other excellent models that are available.

Question 14 – How has modeling improved since the Persian Gulf War?

If we assume that the year 1991 should be used as the base year for comparisons, then there have been major improvements in DOD plume models. Prior to 1991, DOD models were primarily Gaussian plume and puff models from the 1960's, and had not been updated in years. DTRA's HPAC/SCIPUFF model has been developed during the time period since 1991 and enhancements are continuing. Other advanced emergency response models such as the DOE NARAC system, the EPA AERMOD and CALPUFF models, and the NOAA HYSPLIT system have also become available during this time. As a result of problems evident during the Chernobyl release, plume models for radiological releases have been improved and data communications enhanced so that these models can be confidently used in real-time for emergency response. In all of the above cases, the improvements have primarily concerned the improved specifications of wind fields and the improved measurements of wind fields and communication of data. Additional improvements have involved parameterization of wind flow and dispersion in urban areas, using recent field experiments such as the 2000 field study in Salt Lake City.

Further comments -

There is much variability in plume model predictions. I believe that an advanced plume model should be able to demonstrate improved performance over a simple baseline plume dispersion model. This is the same as saying that a new weather forecast model should be able to do better than simple estimates such as climatology or persistence.

There is currently much research underway on development of detailed Computational Fluid Dynamics (CFD) models which can predict the variation of concentrations over small time scales (one second) and over small grid volumes (about 1 m^3). Many persons say that, as computers improve, the CFD models can eventually be practical for emergency response use. These persons stress the CFD models' usefulness in urban scenarios. However, it is not clear whether these models' predictions will be any more accurate than the predictions of models such as HPAC or NARAC, with the problems arising due to the large natural variability of the atmosphere. Personally, I feel that CFD models are overkill, but is possible that I may be proven wrong over the long run. And certainly CFD models are capable of producing impressive detailed color graphics for use by emergency responders. Of course another use of CFD models is as a "data base" for development of simplified parameterizations in model such as HPAC.